CHAPTER 4

Statistical mechanics of DNA melting in confined geometry

In chapter 2, we have studied the migration of polymer under different conditions in presence of geometrical trap. In this chapter we would like to study the equilibrium properties of DNA and the migration in other types of confinement, such as nanopore. There are considerable amount of research works on migration of DNA through a pore having a membrane that differentiates two different kinds of solvent [106–108]. Mycobacterium smegtatis porin A (MSPA) pore is one of the prime examples of biological pores which hold the forte of nano-pore sequencing [109–112]. The size of the pore matches well with the size of nucleobases of DNA, and thus appears to be a good candidate to study the translocation of DNA. The crystal structure of MSPA pore looks similar to a coneshape channel.

In recent years there has been considerable interest in studying conformational changes of DNA in confined environment [113]. The motivation originates to understand DNA replication and transcription in living cell, which is crowded with other biomolecules.

Extensive research has been carried out for past decades on DNA melting due to its significance in denaturation mapping which uses sequence dependence melting as a tool to map the whole DNA in genome [114, 115]. Reisner et. al. performed experiments on melting of flurosense labeled DNA for optical mapping [116]. Marie et. al. showed thermal melting of long DNA ($\sim$ 2 Mb) extracted from chromosome can enhance its readability as well as provide unique details about the repetitive units in sequence[116]. Werner et.al. performed simulations on impact of sequence heterogeneities on melting profile of confined DNA[117]. Reiter-

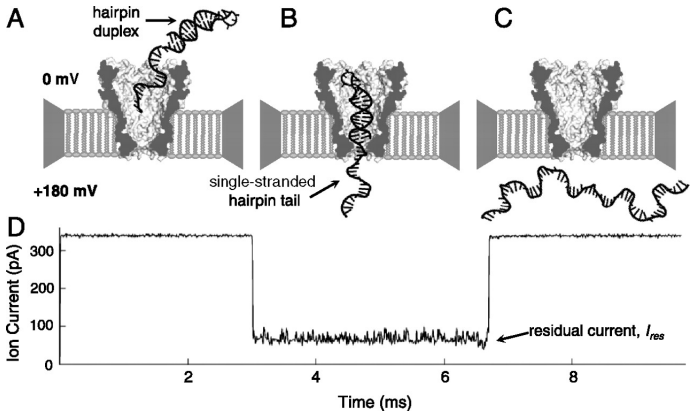

Figure 4.1. Schematic representations of dsDNA translocating through a conical shape pore. Figure A), B), C) shows different stages of translocation. Figure D) represents the associate current distribution during translocation. Whenever the DNA is translocating through pore there is residual current distribution I_{res} which depicts the time of translocation. This figure is taken from ref [110]

Schad et. al. investigated how nanochannel confinement affects the melting temperature of DNA by implementing Poland-Schrega model[118].

DNA is a charged molecule which can be driven through the pore by applying external voltage. This is known as forced translocation and is usually divided into three processes: capture, translocation and ejection from the pore, which makes the whole process complex [119]. Even in the absence of applied force, DNA translocation through pore is a challenging problem (see Fig. 4.1). In this chapter we focus on the unforced translocation of biomolecule through a coneshape channel (mimics the MSPA protein pore Fig. 4.2) which connects two volumes of different solvent quality (say they differ in dielectric constant because of its electrostatic nature). Now due to differing solvent quality there is emergence of chemical potential difference which along with temperature decide the conformation and energetics of biomolecule across the pore. The competition between change in internal energy and conformational entropy of dsDNA in such confined environment needs to be investigated to understand the equilibrium properties of dsDNA. Since, we are interested

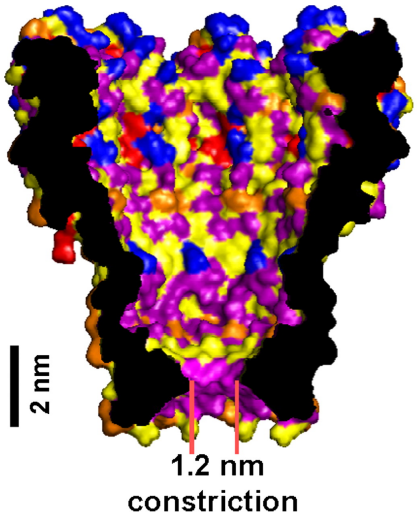

Figure 4.2. Schematic representations of crystal structure of MSPA protein. The figure is taken from the ref [110]

in equilibrium properties we will consider the implicit solvent across the pore. It is a valid assumption as the solvent quality directly affects the base-pairing interactions of dsDNA.

The aim of this present chapter is two fold: first to understand the effect of asymmetry arising due to coneshape channel separating two different kind of solvents on DNA melting. The second aim is to study the free energy landscape of the dsDNA across the pore. For this, we consider MASAW (model A) as described in chapter 1 to model the dsDNA.

In this chapter we briefly describe the model and the method (section 4.1) used to obtain the equilibrium properties of the dsDNA attached to the interface of a coneshape pore, or away from pore in section 4.2. In section 4.3, we study the free energy barrier using Arrhenius kinetics [120] for different solvent conditions. We also estimate the free energy barrier from different approaches by calculating free energy as a function of position co-ordinate x across the pore in section 4.3. To have a better understanding on the asymmetry effect due to coneshape geometry, we compare our results with dsDNA translocating through a pore on the flat surface. In the end of the chapter we add a brief summary of the work in

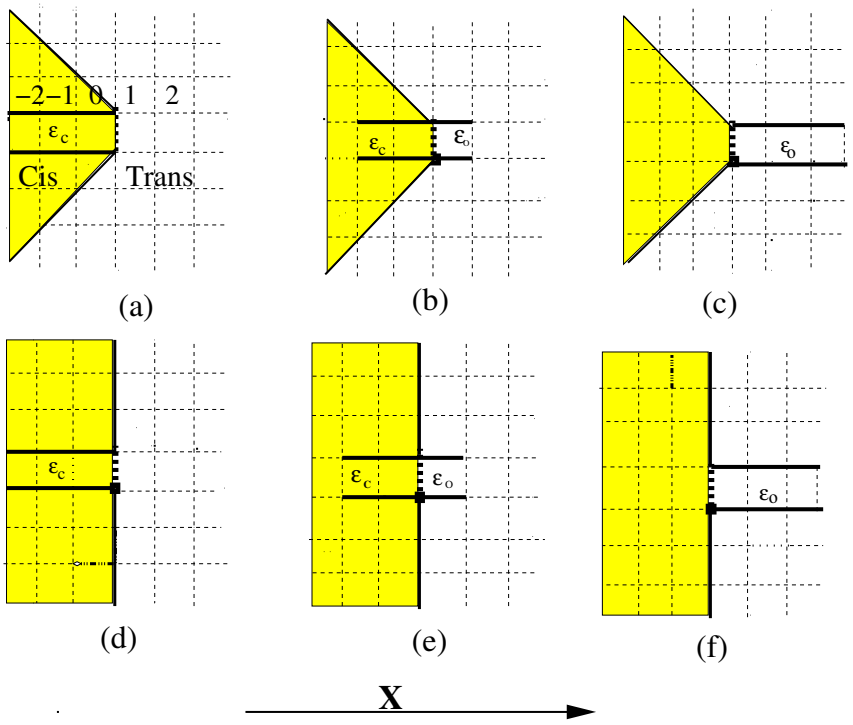

Figure 4.3. Schematic representations of dsDNA attached at different sites across the pore: (i) the cone-shaped channel (a-c), and (ii) the flat channel (d-f). Starting end of the dsDNA is kept fixed and the other end is free to move anywhere except the wall (a and b). To calculate the free energy barrier, we fix the dsDNA chain at varying distances from the interface (say, x). Here, x can be positive, negative and zero (interface). ϵ_c and ϵ_o correspond to the non-native attraction between the complimentary bases of the dsDNA.

section 4.4.

4.1 Model and Method

We consider a cone-shaped channel of fixed pore size of one unit on the square lattice to study the equilibrium properties of dsDNA in the confined geometry. Two walls of the channel (as shown in Fig. 4.3) separate two liquids in such a way that the one side (*cis*) $\approx \frac{1}{4}$ ($\approx \frac{1}{2}$ for flat surface) is available to one type of solvent, while the remaining region $\approx \frac{3}{4}$ is available to solvent outside the cone shaped channel (*trans*). First, we fix one end of dsDNA chain at the interface ($x = 0$) and the other end of

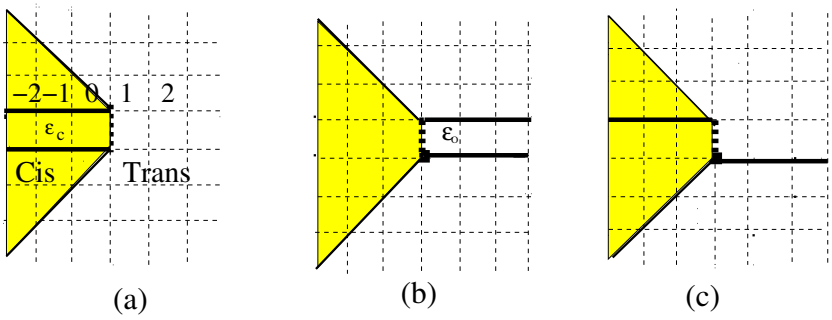

Figure 4.4. Schematic representations of dsDNA chain attached at the interface: (a) dsDNA is completely inside the cone, (b) Completely outside the cone, and (c) One strand is inside the cone, while the other is outside and *vice versa*.

each strand is free to be anywhere Fig. 4.3. Now restrictions are imposed in such a way that the dsDNA chain is not allowed to be on the wall or cross the wall except through the pore. To calculate the free energy barrier across the channel, one end of dsDNA is fixed at varying distances from the interface (say, x=..., -4, -3, -2, -1, 0, 1, 2, 3, 4,...) and the other end is free to move in the region except at the wall.

The thermodynamic properties associated with the DNA melting are obtained from the following partition function which can be written as a sum over all possible Boltzmann weighted configurations of dsDNA:

$$Z^x(\omega, u) = \sum_{(N_{pc}, N_{po})} C_N^x(N_{pc}, N_{po}) u^{N_{pc}} \omega^{N_{po}}. \tag{4.1}$$

Here, N is the chain length of each of the two strands. N_{pc} and N_{po} are the number of base pairs inside and outside the cone-shaped channel, respectively. $u = \exp(-\beta\epsilon_c)$ and $\omega = \exp(-\beta\epsilon_o)$ are the Boltzmann weights associated with base-pair interactions inside (ϵ_c) and outside the pore (ϵ_o), respectively. Here, $\beta = 1/k_BT$. From here onwards we set

$k_B = 1$ and do the analysis in the reduced unit. $C_N^x(N_{pc}, N_{po})$ is the total number of conformations having N_{pc} and N_{po} bound base pairs whose starting point is at distance x from the interface. We consider x to be negative for the *cis* side, and for the *trans* side, x is positive. We use exact enumeration technique to obtain the $C_N^x(N_{pc}, N_{po})$ for $N \leq 14$ monomers.

First, we study melting profile of DNA anchored at the interafce ($x = 0$) of the pore. We employ the ratio method along with the associated Neville table [121, 122] for the extrapolation (the limit $N \to \infty$) to obtain the reduced free energy of the system. The reduced free energy per base pair for $\omega = u$ (*i.e* $\epsilon_c = \epsilon_o$) can be obtained from the following relation

$$G(\omega) == \log \mu(\omega) = \lim_{N\to\infty} \frac{1}{N} \log Z^0(\omega) \tag{4.2}$$

The transition point (T_m) can be obtained from the peak value of $\frac{\partial^2 G}{\partial(\log\omega)^2}$ *i.e* fluctuation in nearest neighbor base pairs. For a given geometrical condition, a dsDNA can be completely inside the cone, or outside the cone, or one strand may be inside the pore, while the other strand remains outside the pore and *vice versa*. For all these cases melting profiles have been obtained using the above protocol.

For finite N, one can calculate exactly the average number of base pairs outside the channel using the following equation:

$$\langle N_{po} \rangle = \frac{1}{Z} \sum_{(N_{pc}, N_{po})} N_{po} C_N(N_{pc}, N_{po}) u^{N_{pc}} \omega^{N_{po}}. \tag{4.3}$$

Similarly, one can calculate $\langle N_{pc} \rangle$ corresponding to base-pairs inside the channel.

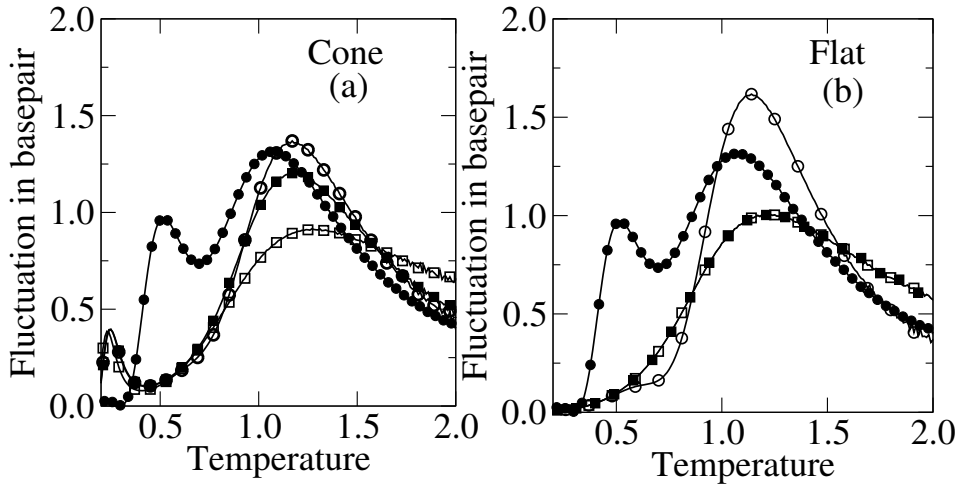

Figure 4.5. Variation of fluctuation in base-pairs with temperature for the cases when DNA is attached at the edge of the pore: Open square represents when both strands are completely inside the cone, whereas solid square represents a case where both strands are completely outside the cone. Open circle stands for the case when both strands can be anywhere. For the sake of comparison, we have shown the melting profile for the case when there is no confinement by solid circles.

4.2 Thermal melting profile of DNA attached at the edge of pore

Melting temperature (T_m) is defined as the temperature where half of the base-pair denatures. Theoretical analysis reveals that T_m is well consistent with the peak of base-pair fluctuation curve [123–126]. We focus our study on the effect of conical confinement on the melting temperature when the dsDNA is anchored at the edge of pore. First we choose the base-pair interactions same (say, same dielectric constant of the solvent) across the interface. In this case chemical potential difference is zero ($\Delta\mu = 0$), and only configurational entropy difference across the pore due to the asymmetry in pore size decides the equilibrium statistics of DNA. For the sake of comparison, we obtain the melting point (T_m) of DNA (under same condition) without any confinement *i.e.* the bulk melting point [126].

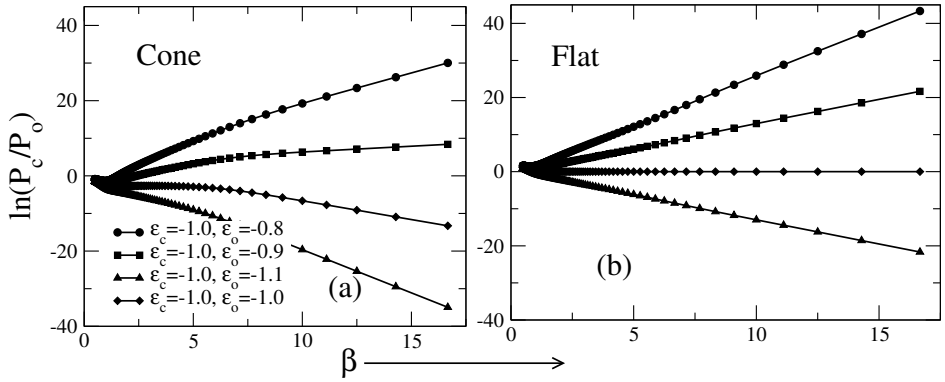

Figure 4.6. (a) Variation of $ln(\frac{P_c}{P_o})$ with β for cone-shaped channel. The slope ΔF gives the free energy barrier; (b) Same as (a), but for flat case.

In general there is an emergence of three scenarios when DNA anchored at pore which is shown in Fig. 4.4. First, we consider the case when the DNA remains in the *cis*-side only. The partition function defined in equation (4.1) will be independent of N_{po} *i.e.* number base pairs remains in the *trans*-side. Following the method described in the previous section we show the melting profile in Fig. 4.5 by open square. Similarly, we obtain the melting profile when the DNA is completely in the *trans*-side. It is evident from the plots that the melting temperature decreases compare to the *cis*-side. This is because of increase in entropy in the *trans*-side. This is also in accordance with the findings of previously obtained results where increasing confinement result in decreasing melting temperature [127]. However, for the flat-shaped channel melting profile overlaps with each other Fig. 4.5 (b) indicating the similar conditions across the interface.

There may be a situation when one strand is in the *cis*-side, while other strand is in the *trans*-side and *vice versa*. Such situation occurs when a pore size is such that it does not allow to pass more than two strands simultaneously [128–130]. For errorless sequencing this is necessary as broad pore size can result stochastic diffusion through the channel. We can calculate the combined partition function which includes all three cases in canonical ensemble. For this case, we find that the melting tem-

perature of dsDNA confined in cone shaped channel is higher compare to the bulk case Fig. 4.5 (a).

One of the challenging issues in such problem is to calculate the free-energy barrier across the interface. It may be noted that the height of the free energy barrier depends on the change in entropy (confinement) and change in solvent conditions across the interface. In present case, this may be calculated from the principle of detailed balance of energy [46]. If P_c and P_o are the probabilities of finding the monomers in the *cis*- and the *trans*-side, respectively, one can write

$$\frac{k_{c-o}}{k_{o-c}} = \frac{P_c}{P_o}, \tag{4.4}$$

where k_{c-o} is the rate coefficient of ejection from the *cis*-side (c) to the *trans*-side (o). The rate coefficient is expected to follow the Arrhenius kinetics [120], $k_{c-o} = k^* \exp[-\beta \Delta F]$, where ΔF denotes the height of the free energy barrier associated with ejection from the *cis*-side to *trans*-side and k^* is a constant. It may be noted that the value of P_c and P_o may be obtained exactly from Eq. 4.4. In Fig. 4.6, we plot the $\ln(\frac{P_c}{P_o})$ with β for different values of ϵ_o, keeping ϵ_c fixed. It is seen from the figure that when β is maximum (at low temperature) the *trans* side is both energetically and entropically favorable for $\epsilon_o = -1.1$, $\epsilon_c = -1.0$. Even for the same solvent situation ($\epsilon_o = -1.0$, $\epsilon_c = -1.0$) entropy in *trans* side dominates over base-pair energy as the slope is negative. Now, when outside is of good solvent quality ($\epsilon_o = -0.9$, $\epsilon_c = -1.0$), then dsDNA is preferred to be found at inside for low tmperature (T=0.2) as the slope is positive. The slope of the curve further rises for the case $\epsilon_o = -0.8$, $\epsilon_c = -1.0$. So, the biopolymer remains inside at low temperature whenever the *cis* is of comparatively poor quality. Here, the entropic contribution is not strong enough to pull the DNA out of the energetically rich side. The major difference in flatshape channel arises for the same solvent case. For $\epsilon_o = -1.0$, $\epsilon_c = -1.0$, both and side are energetically and entropically symmetric to the dsDNA. Due to this, the slope in the

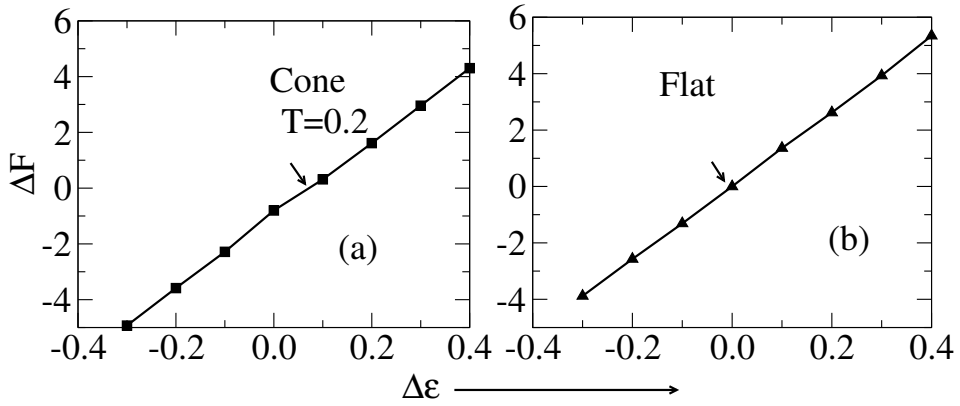

Figure 4.7. (a) The free energy barrier ΔF as a function of $\Delta\epsilon$. The linear dependence is apparent from the plot. (b) same as of (a) but for the flat interface pore. For a cone-shaped channel, it occurs at $\Delta\epsilon \neq 0$, where as for flat pore it occurs at $\Delta\epsilon = 0$. Arrow indicates the value at which the free-energy barrier vanishes.

$\ln\frac{P_c}{P_o}$ is exactly zero. These results are more clear when we plot the slope of the curve which gives the height of free energy barrier *i.e.* ΔF as a function $\Delta\epsilon = \epsilon_c - \epsilon_o$, which shows the linear dependence of slope on $\Delta\epsilon$ in Fig. 4.7. Here, we have kept ϵ_c constant (= -1) and varied ϵ_o systematically, such that $\Delta\epsilon$ varied in the range of -0.3 to 0.3. It is interesting to note that for the flat-channel the slope becomes zero at $\Delta\epsilon = 0$, whereas for the cone-shaped channel, it becomes zero at ≈ 0.1. It is again reinstated that dsDNA will prefer to be inside the *cis* whenever the base pair interaction is higher compared to the *trans*-side ($\epsilon_o = -0.9$, $\epsilon_c = -1.0$). In case of $\epsilon_o = -1.1$, $\epsilon_c = -1.0$ dsDNA comes out from the pore as slopes of these two curves are negative. Now for the flat-shaped channel dsDNA prefers to stay where the base pair interaction is higher. For the same base-pairing interaction (same dielectric constant), the probability of nucleotides being in either sides of flat channel is exactly half as there is no preferred choices.

To have a better understanding of the free energy profile, we plot in Fig. 4.8 the free energy as a function of average number of base pair outside the cone ($\langle N_{po}\rangle$) for different solvent conditions at fixed temperature

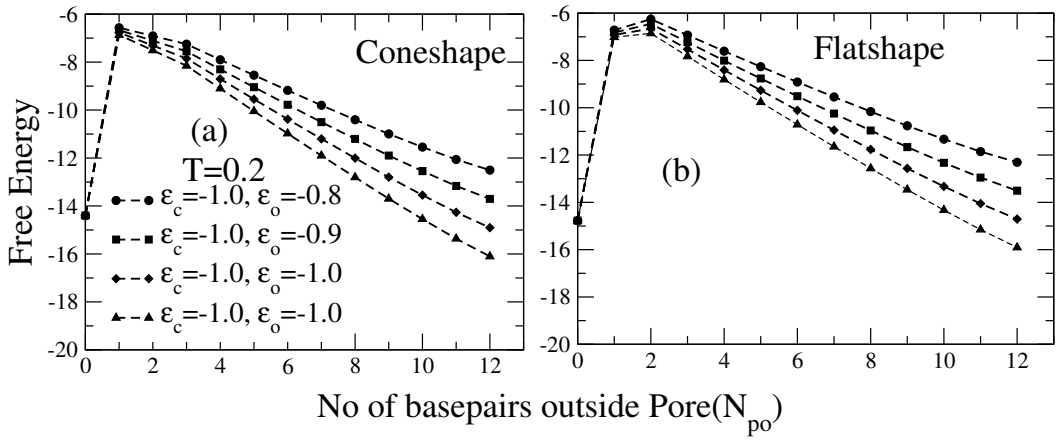

Figure 4.8. Variation of free energy as a function of $\langle N_{po} \rangle$, whose one end is fixed at the edge of (a) the cone-shaped channel. (b) the flat channel. The free energy barrier occurs at maxima of $F(\langle N_{po} \rangle)$.

$T = 0.2$. It has been reported in recent work by Kumar et. al. [131] that for translocation, a minimum number of monomers should be transferred from the *cis*-side to the *trans*-side, otherwise, there is a tendency for the chain to go back again to the *cis*-side. If average number of monomers outside the cone (m_o) exceeds the critical value, because of the downhill, the transfer to *trans*-side becomes more favourable.

We observed here that at low temperature ($T = 0.2$), the maximum of free energy arises at a fixed value of $\langle N_{po} \rangle$, which in turn establishes the fact that a minimum no of base pairs must be transferred to the *trans* side for a successful translocation. It is evident from Fig. 4.8 that for cone shaped channel the critical value of $\langle N_{po} \rangle$ is 1, whereas there is no clear critical value of $\langle N_{po} \rangle$ for flat shape channel. Whenever a single base pair is transferred irrespective of the solvent qualities across the pore, translocation will become easy for the whole chain.

4.3 Melting profile and free-energy landscape of DNA

In order to study how the confinement affects the melting of DNA, we plot the melting temperature (T_m) as a function of different starting positions (x) across the pore in Fig. 4.9. As the chain approaches from *cis*-side to *trans*-side, near the interface the melting temperature first increases and reaches to its maximum at $x \sim -4$ *i.e* in the *cis*-side. This is due to the reason that at bulk the configurational entropy is highest and as the dsDNA approaches the pore number of configurations gets considerably reduced due to the confinement. Now, as the dsDNA advances to a certain extent towards the pore from cis side it has access to the *trans*-side and entropy again starts increasing. So, in the scope of this model there is a minima in terms of configuration entropy at position x=-4 which corresponds to the peak in melting temperature landscape. We observe that the melting temperature decreases if one approaches toward the pore even if the starting position are in the *cis*-side. Once the starting position of the chain enters into the *trans*-side, melting temperature again increases upto a distance $x \sim 2$. This is in accordance with earlier studies where the monomer density near the impenetrable surface was found to be maximum [132]. The peaks in the *cis* and *trans* side also indicate the effect of confinement near interface. Once the starting position is away from the wall of the pore, the melting temperature decreases and approaches to the bulk value [125, 126]. It may be noted that for the flat-shaped channel the peak value of melting temperature is same across the pore. Change in base pairing interactions outside the pore can reduce or enhance the melting temperature with respect to the melting temperature of DNA inside the pore. This has been clearly depicted in Fig. 4.9. It is interesting to note that the qualitative behavior is almost the same for both cases, except the lower melting temperature for flat-shaped channel near the interface.

In order to suppliment the occurance of peaks near the wall, we have

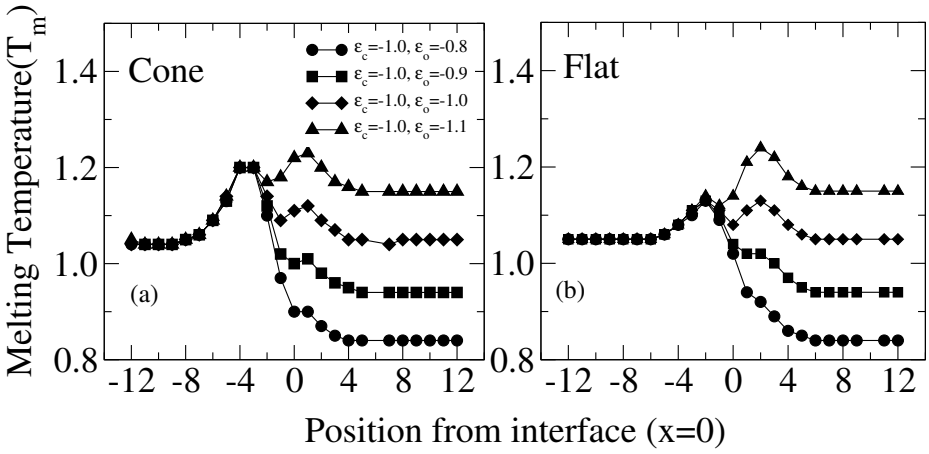

Figure 4.9. Melting profile of DNA across the pore ($x = -12$ to $x = 12$, including zero) for different solvent qualities across the pore: (a) for the cone-shaped channel, and (b) for flat-shaped channel.

plotted the average number of base-pairs outside the cone Fig. 4.10 (a-c) and inside the cone in Fig. 4.10 (d-f) as a function of temperature for different values of ϵ_o at fixed $\epsilon_c = -1$ for different starting points near the pore ($x = -4, -2, -1, 0, 1, 2$). When DNA starts far away from the interface, it will not experience any confinement and, therefore, behaves as if the system is in the bulk. In this case, with the rise of temperature, two strands of DNA go from a compact spiral-like state to a zipped state followed by the melting at a higher temperature. This is evident from the fluctuation curve shown in Fig. 4.5 , and this is also in accordance with previous findings [125, 126]. When base pair interactions are same across the pore ($\epsilon_c = -1$, $\epsilon_o = -1$), for DNA chain starting at $x = -1$ to 1, DNA chain prefers to stay outside the cone in the zipped form at low temperature. With the increase in temperature, $\langle N_{po} \rangle$ decreases, whereas $\langle N_{pc} \rangle$ remains almost negligible. It shows that when DNA chain starts near the interface the system undergoes zipped-unzipped transition outside the pore. The peak corresponds to increase in base-pair density near the interface (at $x \sim -4$ and $x \sim 2$)

When base-pair interaction (inside the cone) is more compare to out-

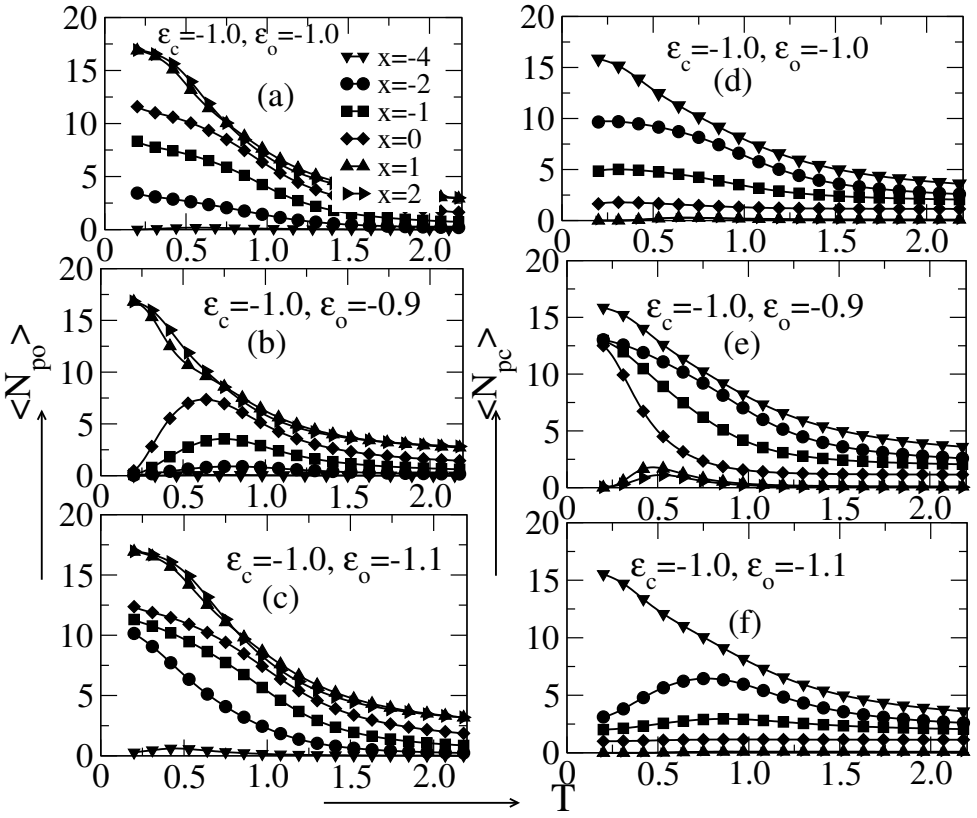

Figure 4.10. Variation of $\langle N_{po} \rangle$ and $\langle N_{pc} \rangle$ with temperature (T) for three sets of interactions and different starting positions of DNA ($x = -4, -2, -1, 0, 1, 2$).

side ($\epsilon_c = -1.0$ and $\epsilon_o = -0.9$) and chain starts at the edge of the interface ($x = 0$), we find that $\langle N_{po} \rangle$ first increases with temperature, reaches a maximum, and then decreases Fig. 4.10 (b) . Instead of going to the unzipped state in the *cis*-side, increase in $\langle N_{po} \rangle$ (corresponding decrease in $\langle N_{pc} \rangle$) indicates that the DNA prefers to be in the zipped state outside the cone. With further increase in temperature, DNA unzip outside the cone Fig. 4.10 (b) . However, when the base pair interaction (outside the pore) is more than that of the inside ($\epsilon_c = -1.0$ and $\epsilon_o = -1.1$), $\langle N_{po} \rangle$ is significantly large for $x = -2, -1, 0$ at low temperature and decreases with further increase in temperature (Fig. 4.10(c)). It is interesting to note that when DNA starts at $x = -2$, instead of unzipping outside the cone, DNA has a tendency to stay inside the *cis* side and unzip with further increase

in temperature which is evident from the increase in $\langle N_{pc} \rangle$ as shown in Fig. 4.10 (f) . This may be due to increase in outside base pair interaction which compel DNA to stay in entropcally favourable *trans*-side.

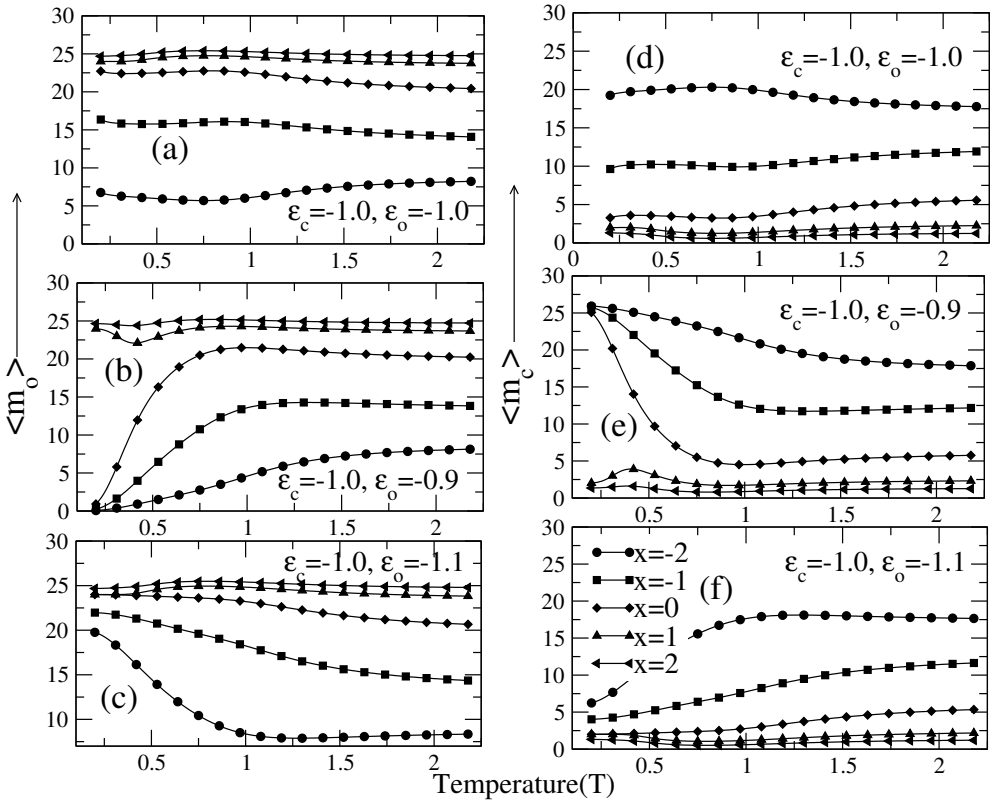

Figure 4.11. shows the variation of $\langle m_o \rangle$ and $\langle m_c \rangle$ with temperature (T) for three sets of interactions and different starting positions of DNA ($x = -4, -2, -1, 0, 1, 2$).

These results are further substantiated by the monomer distribution for different starting positions of the DNA across the pore. For the same solvent condition ($\epsilon_c = -1.0$ and $\epsilon_o = -1.0$) the monomer distribution remain almost constant with temperature for different starting position. From the Fig. 4.11 (a) it is seen that when the dsDNA is fixed at the edge of the pore $x = 0$, due to favorable entropy it remains always outside the pore (zipped at low temperature and denatured at higher temperature). Even for $x = -1$, the average monomer distribution $\langle m_o \rangle$ is higher than

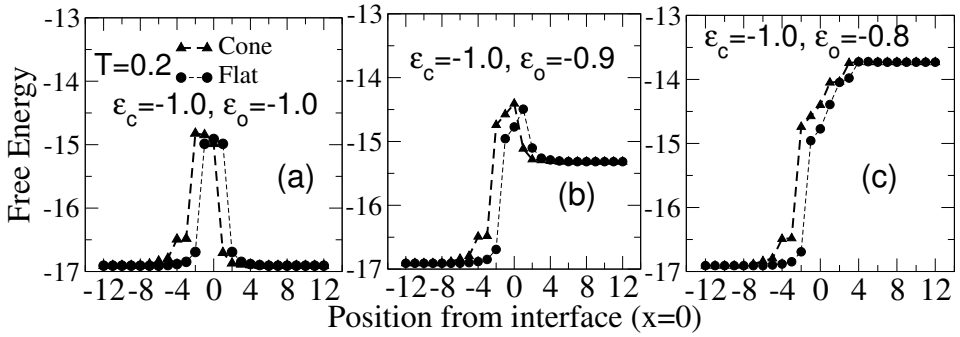

Figure 4.12. Figures (a-c) show the free energy profiles of a DNA chain, whose starting points have been varied systematically from $x = -12$ to 12 for three sets of solvent interactions. (a) If the starting point of chain is far away from the edge, for a given set of interaction, the free energy remains the same for the *cis* and *trans*-side, whereas near the pore it is higher. A rough estimate of the free-energy barrier may be estimated from these plots. This figure also shows the effect of confinement on the free energy arising due to cone-shaped channel and flat channel.(b) For this set of interaction, the difference of free energy between cone-shaped and flat channel vanishes, if the polymer is in the *trans*-side, however, the barrier height increases. (c) Same as Fig. b, but in this case, barrier height increases further. Triangle and circle correspond to the cone-shaped channel and flat-shaped channel, respectively.

the $\langle m_c \rangle$ distribution which states that majority of nucleotides remain outside. Now, interesting phenomenon arises for the case when inside is poor compared to the outside ($\epsilon_c = -1.0$ and $\epsilon_o = -0.9$). It is clear from basepair distribution that at very low temperature dsDNA remains zipped at the cis side for $x = 0$. Now, as temperature increases dsDNA comes out of the pore as entropic contribution favors over internal energy. At high temperature DNA denatures and remains outside the pore even the cis side is energetically favorable (poor solvent side) as $\langle m_o \rangle$ is higher. So, the tug-of-war situation between attractive base-pairing energy and configurational entropy is well controlled by the temperature. Now, for the case when outside is of poor solvent quality DNA always remains outside of the pore at all temperature range for x=0 (starting end is fixed at the interface).

Fig. 4.12 (a-c) show the free-energy profiles of DNA chain, whose

starting points have been varied systematically from $x = -12$ to $x = 12$ at fixed temperature for different sets of interactions. If the starting point of DNA chain is far away from the pore (*cis*-side), then it does not feel any effect of confinement and the free energy of the system remains same as that of the *trans*-side. When the DNA chain approaches the pore of lower conformational entropy and higher internal energy, its free energy increases and becomes maximum near the interface. One can notice the existence of a barrier across the interface. To study the effects of confined geometry we compare the results with flat-shaped channel. It is evident from Fig. 4.12 (a) that the barrier height of cone-shaped channel is relatively higher than the flat channel. The barrier height increases as the base pair interaction in *cis*-side is more than the *trans*-side (Fig. 4.12 b & c).

4.4 Summary

We have studied the melting profile and free energy barrier of a DNA chain anchored to the interface of a coneshape pore using Exact Enumeration technique. Our goal was to study the change in configurational properties of dsDNA arises due to the introduction of conical confinement. The asymmetry in the modelling (coneshape pore) is motivated from the MSPA protein pore which allows dsDNA translocate through it. For uniform chemical potential across the pore (same solvent case), the dsDNA prefers to remain outside of coneshape pore at all temperatures. The comparison with flatshape channel points out that the dsDNA does not have a preferential location unlike in coneshape channel. The simple model presented here captures the essential physics of DNA in the confined geometry. Our studies demonstrated that the confinement arising due to the cone-shaped channel gives rise many interesting properties, which may have potential applications in understanding the biological processes, *e.g.* transcription, translocation etc.